Performing Statistical Analysis on Earned Value Data

Eric Druker, Dan Demangos
Booz Allen Hamilton

Richard Coleman
Northrop Grumman Information Systems

Table Of Contents

Introduction

- Problem Statement
- Performing Statistical Analysis on EVM Data
- Why Statistics are Rarely Used With EVM Data

Introduction: Problem Statement

- Currently, Earned Value Management calculations suffer from several shortcomings that lessen their viability as a cost estimating tool

1. Estimates developed using most EVM equations are subject to tail-chasing whenever the CPI changes throughout the life of a program
   - Tail-chasing is when the EAC for an over running program systematically lags in predicting the overrun, and vice-versa
   - This occurs because these equations are backwards looking in regards to CPI; they lack the ability to predict changes in the CPI looking forward, and fail to perceive trends
   - Tail-chasing is thus inevitable because, as Christiansen wrote: "in most cases, the cumulative CPI only worsens as a contract proceeds to completion."[1]
2. Since the traditional EVM equations are simple algebra, and not based on statistical analysis, estimates developed using them are not unbiased, testable or defensible
   - Bias is the difference between the true value of an estimate and the prediction using the estimator
   - Testable estimates are those which can be subjected to decisions based on measures of statistical significance
3. Quantitative cost risk analysis can not be performed on EVM data without subjective inputs

[1]Christensen, David S (1994, Spring). "Using Performance Indices to Evaluate the Estimate At Completion." *Journal of Cost Analysis and Management, pp 17-24.*

Introduction: Performing Statistical Analysis on EVM Data

- Performing statistical analysis on EVM data solves all of the aforementioned shortcomings
  1. EACs developed using statistics include a forecast for the final CPI and thus are not subject to tail-chasing
  2. EACs developed using statistics are based on historical data, and are therefore testable and defensible
     - Statistical significance can be used to defend the estimate
  3. Statistical methods will produce unbiased estimates that include the uncertainty measures needed for risk analysis
  4. Statistical methodologies can be applied alongside traditional earned value methods and easily incorporated into the EVM process
     - They provide an independent cross-check of the calculated estimates
     - Once the statistical analysis has been performed the first time, it can be updated with very little recurring effort

- Although not discussed in this paper, similar methods can be applied to the SPI to develop statistically based schedule estimates using EVM data

Druker_Eric@bah.com

Introduction: Why Statistics are Rarely Used With EVM Data

- A pre-requisite for just about any defensible cost estimate, statistical techniques have yet to be widely applied to EVM data for various reasons
  - EVM traditionally falls within the realm of program management or financial controls, not within the realm of cost analysis
    - EVM was developed as a program management technique for measuring progress in an objective manner
  - From a cost estimators perspective, it is difficult to acquire the data needed to perform statistical EVM analysis
    - There aren't many databases dedicated to historical EVM data
    - Data gathering/normalization is often the most time consuming part of statistical analysis
  - The techniques needed to perform statistical analysis on EVM data can be complicated, especially when there are events such as rebaselining involved
  - Patterns within EVM data are generally not obvious just by looking at trends on a scatter plot

- Despite the difficulties in applying statistical analysis techniques to EVM data, the ability to produce defensible, unbiased estimates that include risk analysis is well worth the effort

Druker_Eric@bah.com

Performing Statistical Analysis on EVM Data

Druker_Eric@bah.com

Performing Statistical Analysis on EVM Data: Goals

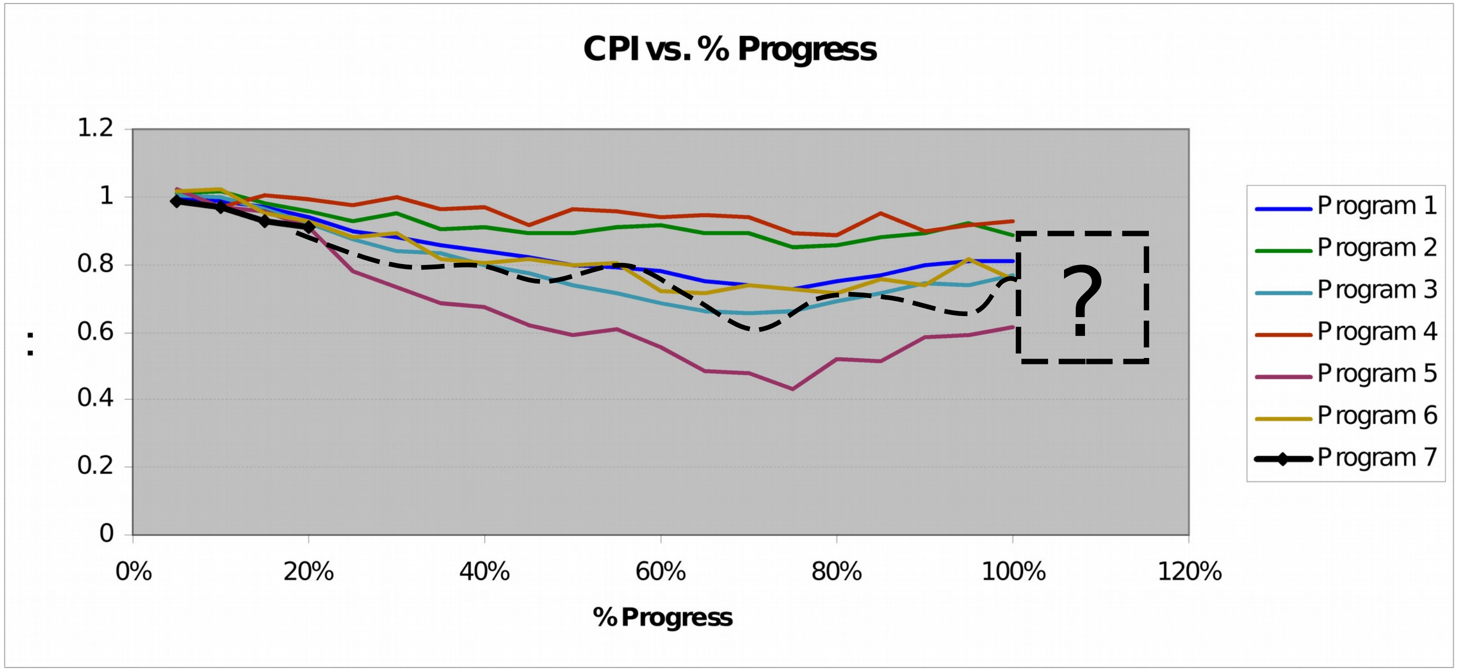
- The theory behind statistical EVM analysis is that programs of a similar nature, or performed by a similar contractor, can be used as a basis to project patterns in the CPI over time
  - Example: For ship production programs, the cost of 1% of progress rises (and thus the CPI drops) over time
    - This occurs as ships move from the shop, to the blocks, to the water, and, e.g., workers move from welding at their feet to welding above their heads
  - Looking only at the current, or average, CPI, estimates for these ship production programs would always tail-chase

- The results of this analysis provides program managers and decision makers with:
  - An EAC that is historically based, unbiased, testable and defensible
    - Testable refers to the ability to apply statistical significance to a relationship
  - The statistical uncertainty around the EAC for use in risk analysis and portfolio management

- An example using representative data follows on the next several slides

Druker_Eric@bah.com

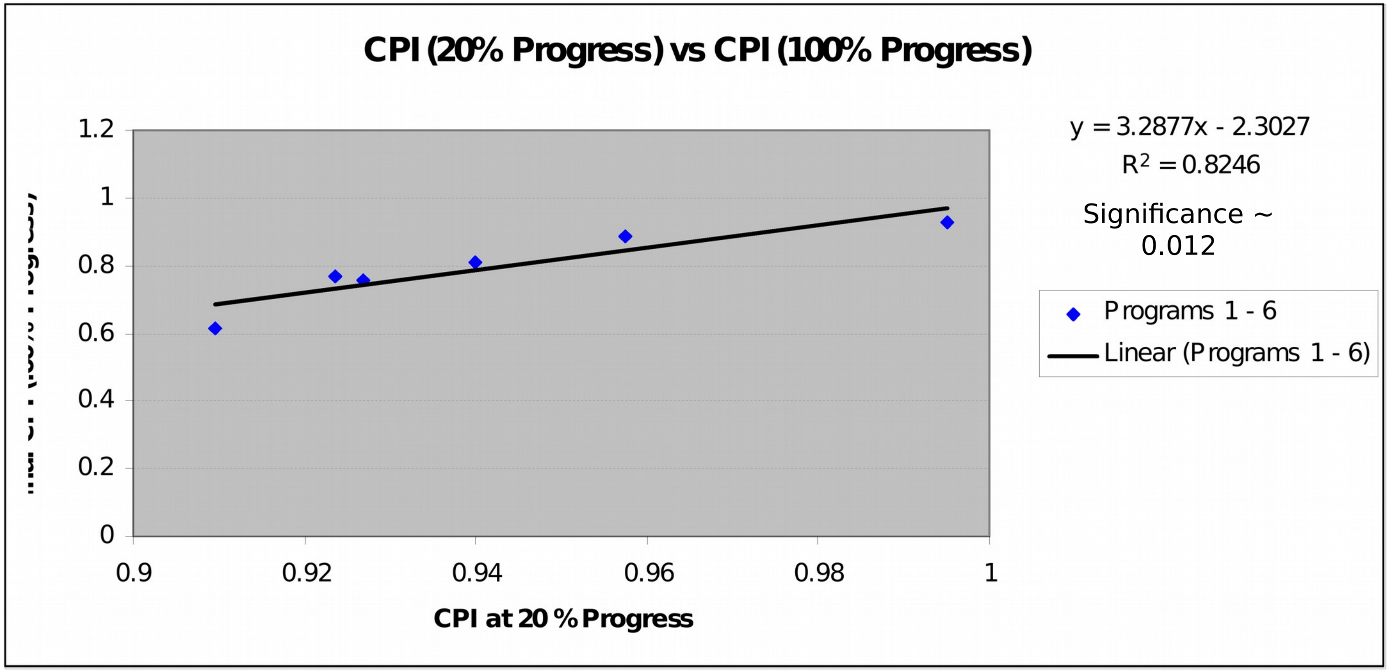
Performing Statistical Analysis on EVM Data: Example



Program 7	
BCWP	$ 20
BAC	$ 100
% Progress	20%
ACWP	$ 22
CPI	0.91

- The above graph shows the CPI over time vs. % reported progress for 7 different programs
  - Examining the lines, it is not apparent that there is a trend that would yield any applications to the in-progress program (Program 7)
- Data from Program 7's latest EVM report is on the right

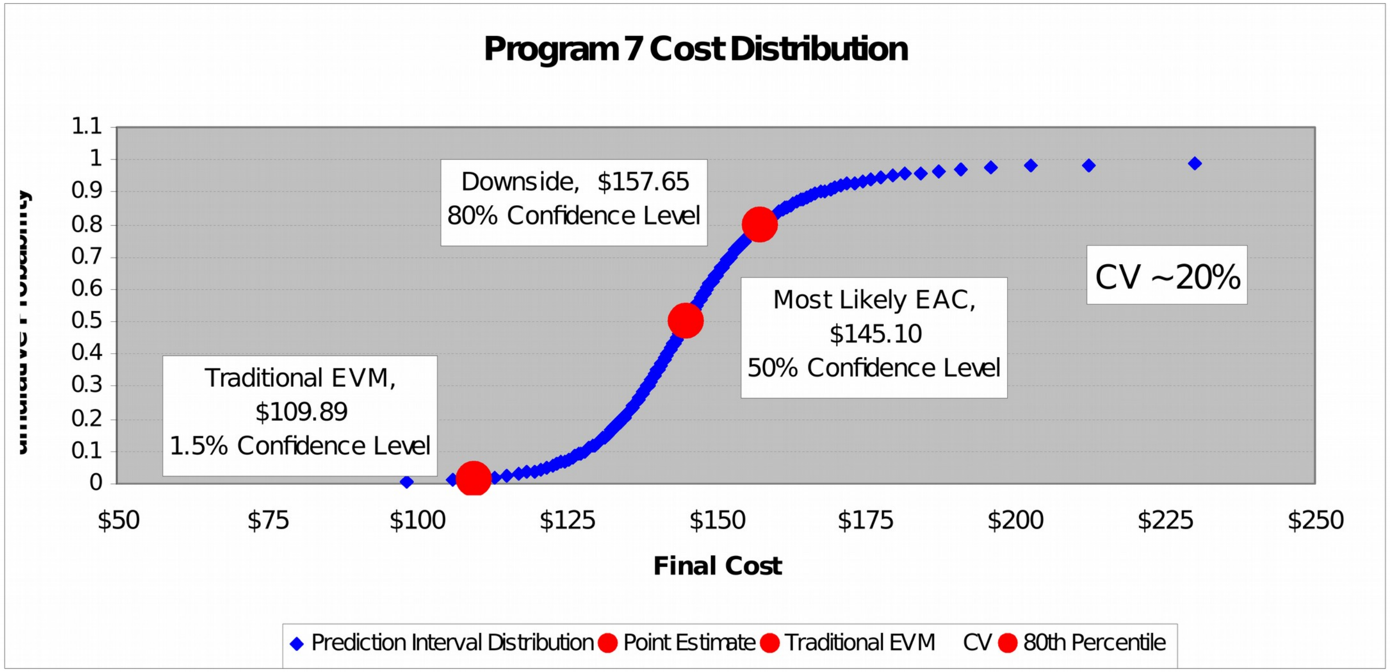
Performing Statistical Analysis on EVM Data: Example



Program 7	
BCWP	$ 20
BAC	$ 100
% Progress	20%
ACWP	$ 22
CPI	0.91

- With a closer look at the data, it is revealed that there is a significant relationship between a program’s CPI at 20% progress and its final CPI
  - This implies that a program’s CPI at 20% progress can be used to estimate its final CPI and thus its EAC
- This relationship (and others like it) will be used to develop a new estimate for Program 7

Performing Statistical Analysis on EVM Data: Example



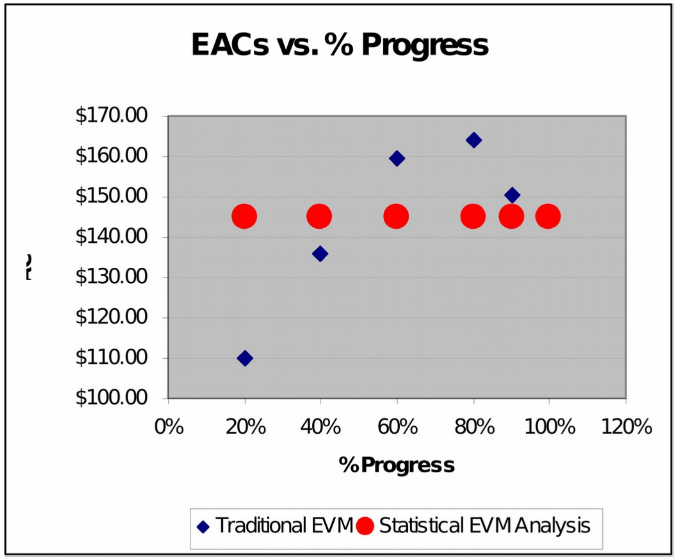
Program 7	
BCWP	$ 20
BAC	$ 100
% Progress	20%
ACWP	$ 22
CPI	0.91
Traditional EVM	
EAC	$ 109.89
Statistical EVM Analysis	
Predicted Final CPI	0.69
EAC	**$ 145.11**

- Using the knowledge gained from the regression analysis, a predicted final CPI of 0.69 (rather than the current reported CPI of 0.91) is applied to the BAC
  - This EAC differs dramatically from that produced using traditional EVM
  - More importantly, it is statistically significant and unbiased
- Because statistics were used to develop the estimate, the risk curve is a byproduct of the estimate

Booz | Allen | Hamilton

Performing Statistical Analysis on EVM Data: Example

Program 7	% Progress					
	20%	40%	60%	80%	90%	100%
BCWP	$ 20	$ 40	$ 60	$ 80	$ 90	$ 100
BAC	$ 100	$ 100	$ 100	$ 100	$ 100	$ 100
% Progress	20%	40%	60%	80%	90%	100%
ACWP	$ 22	$ 54	$ 96	$ 131	$ 136	$ 145
CPI	0.91	0.74	0.63	0.61	0.66	0.69
Traditional EVM						
EAC	$ 109.89	$ 135.94	$ 159.36	$ 164.05	$ 150.68	$ 145.11
Statistical EVM Analysis						
Predicted Final CPI	0.69	0.69	0.69	0.69	0.69	0.69
EAC	**$ 145.11**	**$ 145.11**	**$ 145.11**	**$ 145.11**	**$ 145.11**	**$ 145.11**



- In the chart above, EACs developed using the gold card equations change with each data drop
  - This is an example of EVM producing biased estimates
- Statistical analysis uncovers that the CPI exhibits *predictable* trends over time and thus some changes in the CPI over time can be anticipated
  - Since these shifts in the CPI are predictable, the data can be normalized to yield an unbiased EAC that will not change so long as Program 7 behaves similarly to the historical programs

Performing Statistical Analysis on EVM Data: Data Requirements

- This analysis requires EVM data from completed programs of a similar nature
  - Programs performed by the same contractor as is performing the work in question
  - Programs that would be considered close enough an analogy to include in a CER
- Examples of progressing data:
  - Earned value reports
  - Dated cost reports with an estimated completion date
    - Any data that allows a measure of progress to be developed will work (ex: percent of estimated schedule, percent of final schedule, BCWP/BAC, milestones such as PDR, CDR, etc.)
    - The best form of data would be a measure such as first flight or launch, that is a dependable measure of progress
- The most difficult step in this method is not data collection but data analysis
  - Analysis tools such as dummy variables can be used to handle re-baselinings within the data

Performing Statistical Analysis on EVM Data: The Process

- The aforementioned techniques can be easily incorporated to fit within the EVM process
- Due to the comparably high start-up cost for developing statistically-based EVM estimates (generally 1-3 weeks *after* the collection of historical data is complete), these methods are best applied when there is low confidence in the currently available estimates
  - This could be due to the calculated EAC demonstrating tail-chasing, if there is significant variance between the grassroots estimate and the calculated EAC
- Once the statistically-based estimate is available, it provides an independent crosscheck of the available estimates
- Once the statistical analysis is complete, the recurring cost to update the estimate is minimal (4 hours – 1 day)
  - Updating the estimate may not be needed if it verifies the calculated EAC
- The following slides will show the success of this method when applied to an actual program

A Real World Example: Progress Based EACs

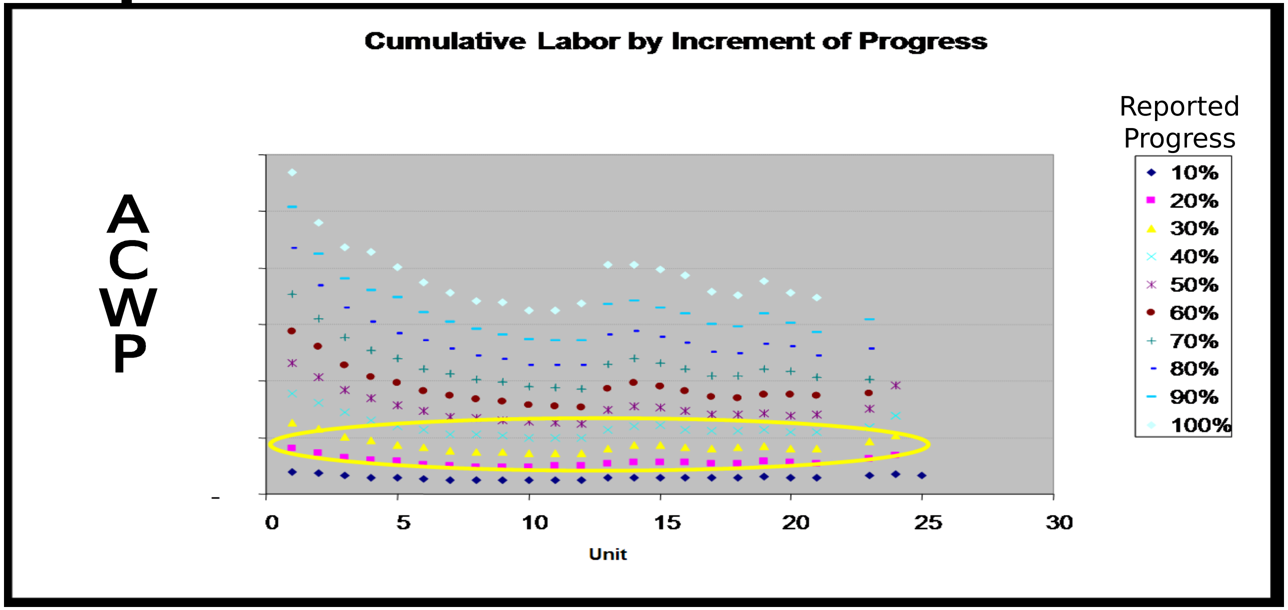
- From the paper: *Ending the EAC Tail-Chase: An Unbiased EAC Predictor Using Progress Metrics*; Druker, Eric, Coleman, Richard, Boyadjis, Elisabeth, Jaekle, Jeffrey, SCEA Conference, June 2006, New Orleans, LA

Introduction

- A client was facing a two-fold problem in estimating production units at their facility
  - Estimates developed using EVM were found to tail-chase and were viewed with wide skepticism by their government client
    - By tail-chase it is meant that by the time an EAC was reported, the latest EVM metrics would already yield an increase above and beyond that EAC
  - A natural disaster had occurred at the production facility causing a sharp and prolonged decrease in productivity
- The PM for one of the programs at this facility reached out to see if there was a way to produce more accurate and defensible estimates than currently available
- The resulting analysis represented the author's first experience with performing statistical analysis on EVM data
  - This specific implementation is known as the Progress-Based EAC method
- This analysis differs from that in the previous example in that the final cost was regressed against ACWPs at various progress points
  - As opposed to the final CPI being regressed against the CPI at various progress points

Booz | Allen | Hamilton

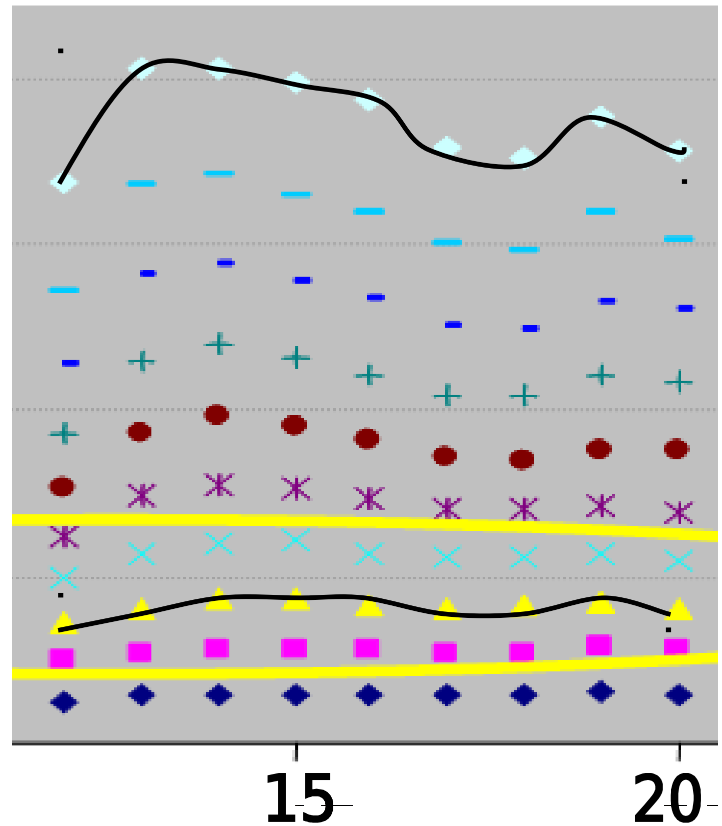
The Key Graphic



- As-reported EVM data was gathered for all units of the same type being estimated that had been produced at the facility
- The ACWP at intervals of 10% progress was scatter plotted on a chart to see if any patterns were visible
- It became immediately apparent that the pattern in the points representing the final cost of each unit became visible as early as 30% of progress

The Key Graphic Continued

- The graph to the right focuses in on units 12 through 20, when the facility experienced unexplained cost growth on many of their units
  - In all cases, this growth was not recognized till the unit was significantly along in its production cycle
- From this graph it is apparent that had the facility compared the ACWP of any two units at equal percent progresses, they would have been able to predict at least relative cost growth
- This chart led to regression analysis being performed on the EVM data
- Could the final cost of a unit be predicted knowing only its ACWP at a certain percent progress?



Regression Results

- At each 10% increment of reported progress, the final cost was regressed against the ACWP
- At 20%, the fist significant regression was found
  - With an *unbiased* error of 4%
- Conclusion: By 20% progress, the facility could predict the cost of any unit, unbiased, ±4%
  - The further along the unit, the less the error

SUMMARY OUTPUT

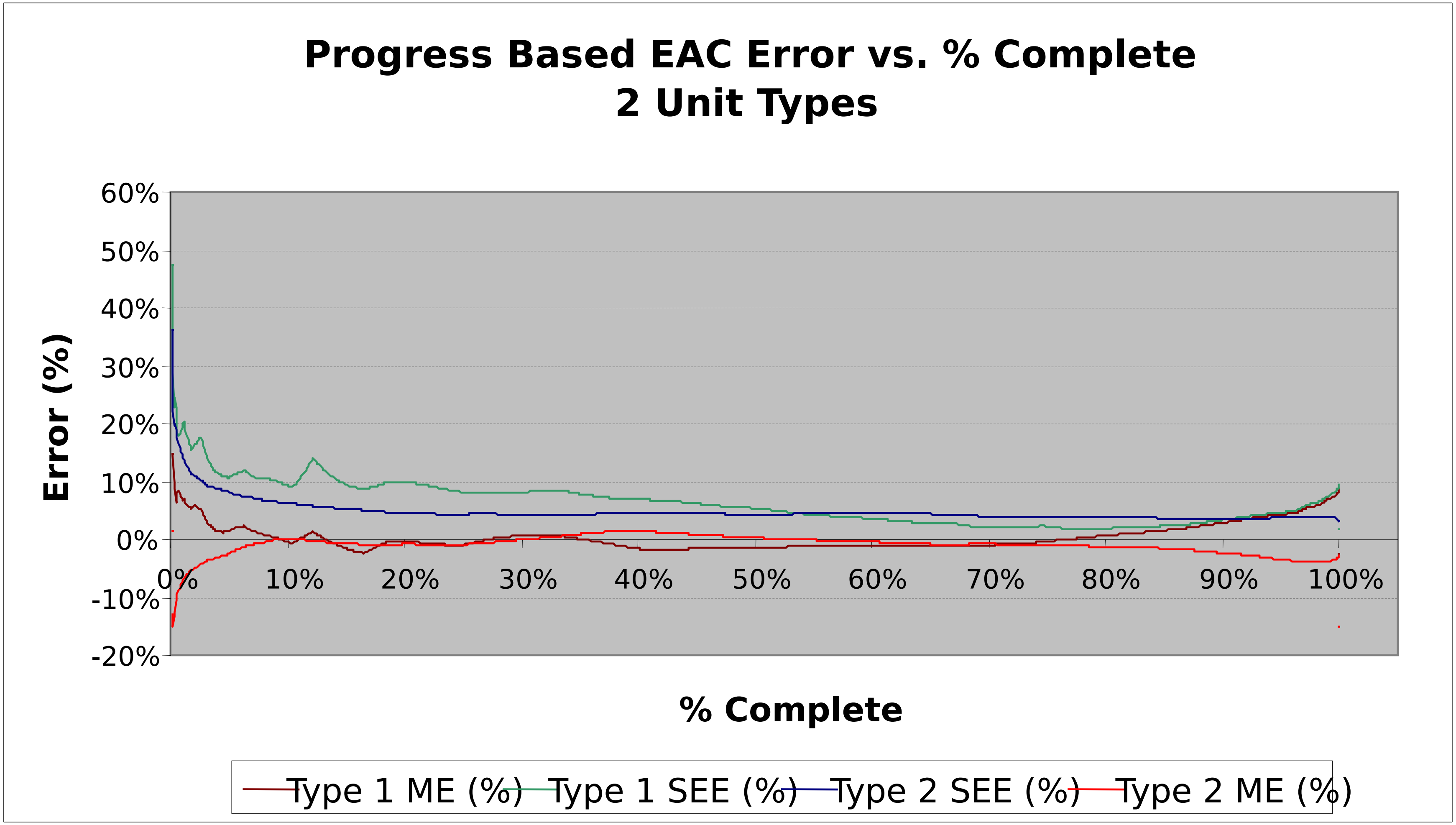
Regression Statistics	
Multiple R	0.956210345
R Square	0.914338224
Adjusted R Square	0.90982971
Standard Error	173979.0514
Observations	21

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	6.13857E+12	6.139E+12	202.80255	1.36728E-11
Residual	19	5.75105E+11	3.027E+10		
Total	20	6.71368E+12			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	152908.7692	262941.1092	0.5815324	0.5677177	-397433.2962	703250.834	-397433.296	703250.834
20%	6.610824914	0.464214768	14.240876	1.367E-11	5.63921224	7.58243759	5.63921224	7.58243759

Regression Results: Error Tracking



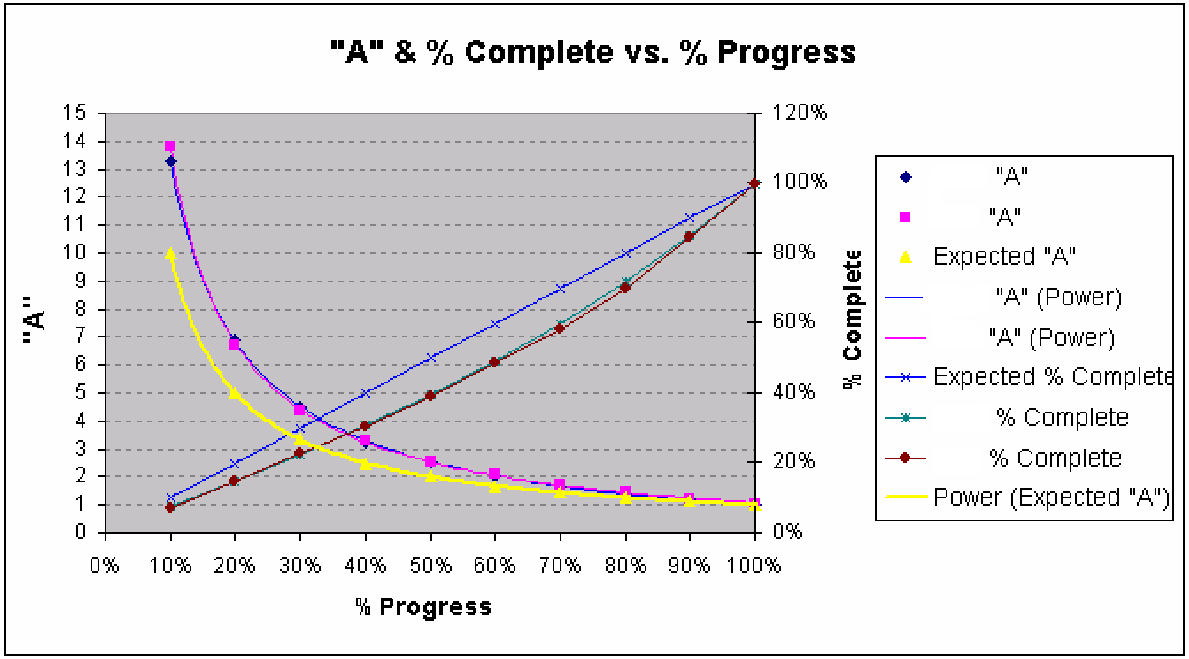
Booz | Allen | Hamilton

Regression Analysis Continued

- With the success of the regression analysis, further work was done to gain more insights
- The next step was to perform a "regression of regressions":
- Each of the previous regressions was of the form: $\text{Final Cost} = A * ACWP_{\%\ Progress} + C$
  - After taking a look at the results, the intercept C was removed from the regression to produce the equation: $\text{Final Cost} = A * ACWP_{\%\ Progress}$
  - *A* represents a "multiplier" that is used to extract the final cost of any unit from an ACWP
  - *1/A* represents the true percent progress in terms of cost
  - C was removed because it was unstable and degraded the utility of the model
    - When *C* was removed the other terms proved sufficiently stable
- With the regressions complete, the *A* term was charted against its associated % reported progress
- These plots were developed for two types of units with different schedules, costs and physical parameters
  - The lines representing the A multiplier for the two types of units were found to be the exact same

Druker_Eric@bah.com

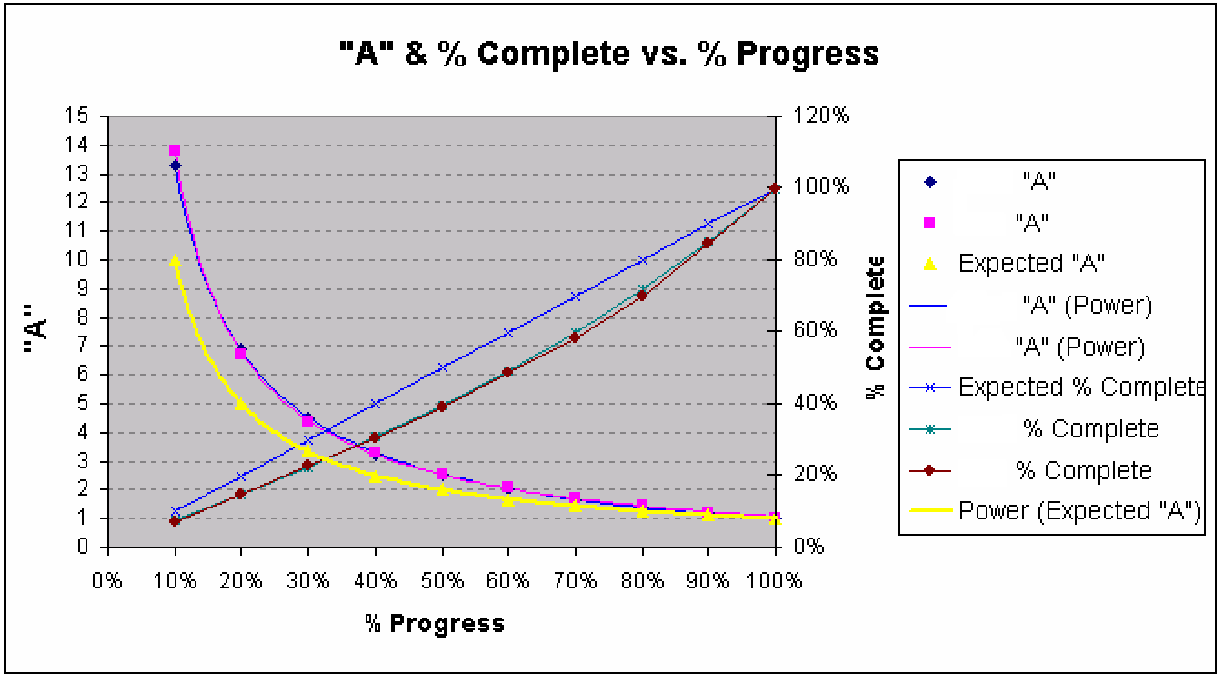
Regression Analysis Continued



- Several breakthrough insights were gained through the above graph

1. As the % Complete (in terms of cost) vs. % Reported Progress line is non-linear, the facility's EACs (using traditional EVM) must tail-chase as the CPI is always degrading
2. The *A* multiplier for both types of units produced by the facility follow the same curve meaning the analysis can be used to estimate units of types not included in the data
3. Each % progress costs progressively more as the unit moves along in production

Estimating Final Cost



- To estimate the final cost of a unit, the *A* multiplier for the current % progress was found from the chart above
  - *A* was then applied to the current ACWP to find the EAC

- For example, an ACWP of $50 at 10% progress would yield an estimate of: $50 * 13.2 = **$690**

Implications

- Since the multiplier lines for two different programs overlay each other, the facility's progress points are standard across unit type and directly related to cost
  - This implied that the method could be applied to any unit produced by the facility, even those that were not a part of the historical analysis
  - This was proved to be true over the next two years
- As the cost per 1% progress rises throughout construction, traditional EVM would never produce an accurate EAC
  - The degrading CPI would lead to consistent tail-chasing
  - This degradation however is predictable a-priori, which is why the method works
- The multiplier curves can be used to predict the ACWP at a future % reported progress
  - Comparing the actual ACWP to this provides a method by which productivity can be monitored

Summary

- This method is a wholly-data-based method of EAC projection that relies upon Progress-and-MH data alone. The model is
  - Able to project EACs for all unit types at the facility within about 2% - 5% after about the 20% progress point
  - Able to work incrementally projecting work remaining given MH
  - Able to include uncertainty with the estimate because it is statistically based
  - Unbiased – the error is symmetric … specifically, it does <u>not</u> result in a tail chase
- In the case of short term effects, the model, because it is progress based, is able to separate out specific effects such as additional costs due to a fire or other exogenous event for units that were at least 20% complete before the event
  - This "effect cost" is obtained by subtracting the as-would-have-been cost from the actual end cost
- In the case of long-term effects, because of its incremental ability, the model is able to add actuals up to an event, and, since it can predict ETC after any post-event increment of about 20% of progress has occurred, can predict ETCs after the event.

Since the Analysis…

- The previous was nothing short of a revelation for the client, who had programs that had experienced multiple rebaselinings
  - To date, the method has correctly estimated the final cost of all 4 units it has been applied to
  - Midway through the production effort of one of these units (in 2006), the Progress-Based EACs method forecasted 60% cost growth in the final cost
  - This cost growth was predicted prior to latest program estimate recognizing **a single dollar of cost risk**
    - After significant resistance, it took a full 2 years (2008) before the program team recognized that 60% cost growth was even feasible
    - It took another 6 months (2009) before the program team recognized that 60% cost growth was, in fact, accurate

- Following this success, the method was expanded
  - This analysis is performed on all in-progress programs and the results are presented to executive management regularly
  - The method is also used to monitor productivity on all in-progress programs

Conclusion

Conclusion

- Performing statistical analysis on EVM data provides an invaluable capability in that:
  - CPI forecasts can be developed, thus avoiding the problem of tail-chasing when estimates are developed using only backwards looking equations
  - The EACs developed using statistical methods are unbiased, testable, and defensible
  - The uncertainty in the estimate, for use in risk analysis, is automatically included with statistically based EACs
  - The analysis can be incorporated into the EVM process to provide a third data point in addition to the calculated EAC and grassroots estimate

- Despite the utility of methods such as these, there are still hurdles to overcome before they can be widely implemented
  - EVM data from completed programs must be compiled and provided to cost estimators
  - Cost estimators must be involved in the EVM process

Druker_Eric@bah.com